* Basics of RDD like which is transformation which is action/base RDD/which initiates execution of linage graph **- 4-5** questions
* Given a python data like:

**a=Set([“Ned”,”Sant”,”Amit”,”Naive”,…….])**

* + b=parallelize(a);
  + **Got 2 qns** on the same data:

b.keyBy(x=>x[0]).groupBy().distinct().map(some transformation) and asked for output

Data set is same but don’t remember qn properly and there were some transformations similar to above one

* 3 qns on shuffles and memory reduction like given the set of code asked which code provides less shuffle or mor-e shuffle or takes less /more memory
* Qn on Hadoop reducer is used in

reduceByKey or CombineByKey…

* Broadcast variable network usage with increasing Nodes by ‘N’

Options were:

* + - * 1. Linearly increases
        2. Varies with O(Nlog(N))
        3. Sub linearly increases
        4. No change
* 2 qns on conversion to Scala to java n python conversion
* Java wordcount code
* Accumulator qn like created

Val acc=sc.accumulator(0)

a=[1,2,3,4]

b=sc.parallelize(a)

c=b.map(x=>acc.value=x)

print(acc.value)

and asked like what is output of above operation given 2 exception options 2 normal output options

* def xyz(x,a,elem):

a[elem]+=1

return a[elem]

a=[“foo”=2,”bar”=3]

b=sc.paralllelize**([“Ned”,”Sant”,”Amit”,”Naive”,…….]**)

c=b.map(x=>xyz(x,a,”foo”))

asked for output

* Serialiazable exception qn : nt a solution to resolve exeption

Options were like set SPARK\_LOCAL\_HOST=127.0.0.1 / change class to serializable / SPARK\_LOCAL\_IP=127.0.0.1

* MLib given kmeans code completely and gave 4 options informing about what is happening in code and asked which option is not true
* Graphx similar example given in our training changing only input data asked what is output
* 2qns --Given SQL query and equivalent of it python/scala

Like sql.where().select

Or sql.select.where or groupby/join etc

* Given a Spark SQL query and asked like which rdd operation provide similar output
* Which window operation is not valid,and batch interval is 4,options like;

.window(8,4)

.window(8)

.window(8,2) and window(4,4)

* Which operation required checkpoint on streaming